This is a digital copy of a book that was preserved for generations on library shelves before it was carefully scanned by Google as part of a project to make the world's books discoverable online.

It has survived long enough for the copyright to expire and the book to enter the public domain. A public domain book is one that was never subject to copyright or whose legal copyright term has expired. Whether a book is in the public domain may vary country to country. Public domain books are our gateways to the past, representing a wealth of history, culture and knowledge that's often difficult to discover.

Marks, notations and other marginalia present in the original volume will appear in this file - a reminder of this book's long journey from the publisher to a library and finally to you.

Usage guidelines

Google is proud to partner with libraries to digitize public domain materials and make them widely accessible. Public domain books belong to the public and we are merely their custodians. Nevertheless, this work is expensive, so in order to keep providing this resource, we have taken steps to prevent abuse by commercial parties, including placing technical restrictions on automated querying.

We also ask that you:

+ *Make non-commercial use of the files* We designed Google Book Search for use by individuals, and we request that you use these files for personal, non-commercial purposes.

+ *Refrain from automated querying* Do not send automated queries of any sort to Google's system: If you are conducting research on machine translation, optical character recognition or other areas where access to a large amount of text is helpful, please contact us. We encourage the use of public domain materials for these purposes and may be able to help.

+ *Maintain attribution* The Google "watermark" you see on each file is essential for informing people about this project and helping them find additional materials through Google Book Search. Please do not remove it.

+ *Keep it legal* Whatever your use, remember that you are responsible for ensuring that what you are doing is legal. Do not assume that just because we believe a book is in the public domain for users in the United States, that the work is also in the public domain for users in other countries. Whether a book is still in copyright varies from country to country, and we can't offer guidance on whether any specific use of any specific book is allowed. Please do not assume that a book's appearance in Google Book Search means it can be used in any manner anywhere in the world. Copyright infringement liability can be quite severe.

About Google Book Search

Google's mission is to organize the world's information and to make it universally accessible and useful. Google Book Search helps readers discover the world's books while helping authors and publishers reach new audiences. You can search through the full text of this book on the web at `http://books.google.com/`

HARVARD UNIVERSITY

LIBRARY OF THE

GRADUATE SCHOOL
OF EDUCATION

TEXTBOOK SELECTION

TEXTBOOK SELECTION

BY

R. H. FRANZEN

DIRECTOR OF RESEARCH, DES MOINES PUBLIC SCHOOLS

AND

F. B. KNIGHT

STATE UNIVERSITY OF IOWA

With an Introduction by
ERNEST HORN

BALTIMORE, MD.
WARWICK & YORK, INC.
1922

HARVARD UNIVERSITY
GRADUATE SCHOOL OF EDUCATION
MONROE C. GUTMAN LIBRARY

Dec. 15. 1922
HARVARD UNIVERSITY
LIBRARY OF GRADUATE SCHOOL
OF EDUCATION
LB 3045
.F7

COPYRIGHT, 1922, BY
WARWICK & YORK, INC.

THE MAPLE PRESS - YORK PA

INTRODUCTION

It is becoming increasingly clear in the minds of those who face the practical problems of improving instruction, that text books usually determine the success or failure of any educational method. And yet the various problems involved in text book making have been attacked mainly through the hazardous and wasteful method of trial and error, or by the method of rough judgments. Recently, however, an inclination to regard the selection of text books as a matter for scientific study has become prevalent. Every study dealing with students or social values or economy in learning has been promptly seized upon by authors or publishers either to guide the construction of new books or to justify books already published.

Studies of social utility in the various school subjects have been particularly effective in changing the content of text books. This is easily seen in the reconstruction of books in spelling, and in the reorganization of the texts in arithmetic. It is now practically impossible to sell spelling books which do

not at least claim to be based upon the recent scientific studies of the vocabulary of correspondence.

There is also rapid response to varying types of educational theory. Both the subject matter and the organization of books are effected as the propaganda for new theories becomes wide spread. However, both publishers and superintendents are coming more and more to distrust mere philosophizing as a method of determining the organization of school books; the publisher because these philosophical theories are often transitory; the superintendents because books based upon these theories are often found to be unsatisfactory. It is becoming increasingly clear that all theories for the construction of text books must line up with the criteria now available from the scientific appraisals of social needs, and with the criteria available from the experimental studies in the economy of learning.

These studies in the economy of learning, including the study of performance tests, are of the deepest significance. Some data are now available in almost every field and kind of learning. They range all the way from those learning the more elusive problems of determining tastes and interests to the somewhat simpler problems involved in determining practice exercises for specific performances such as hand writing. The two studies reported in this book are excellent examples of the utilization of scientific

techniques in determining student reactions to text book materials.

At a time when a new emphasis upon motivation is being made, Dr. Knight's study of students' choices in literature is very timely. We have only begun to appreciate the importance of utilizing the method of judgments of choices under properly controlled conditions. Certain aspects of learning are particularly in need of just such appraisals. It is fortunate to have an example which does not end merely as a statement of a general principle but in a helpful, concrete recommendation which can be immediately utilized by all teachers of high school English.

It was inevitable from the first that the various studies of economy in silent reading would react upon the construction of text books. Dr. Franzen's investigation was made in the practical emergency of text book selection. It illustrates what may be expected more and more in the way of substituting data for opinion in the solution of the problems involved in text book adoptions. There is reason to believe that in the next few years we may expect much experimental activity along these lines.

Both of the studies suggest the interesting possibility of subjecting manuscripts to some such experimentation before printing. It would seem that authors owe this to the publishers, and both authors and publishers owe this to the public. Of course, scientific

studies take time and are expensive. It is not unreasonable to expect that the profession of teaching is now on such a basis that the rewards for taking such pains will make the research worth while.

ERNEST HORN.

CONTENTS

PAGE

INTRODUCTION 5

CHAPTER I

THE CRITERIA FOR THE SELECTION OF TEXTBOOKS . . . 11

The Factor of Interest.
The Factor of Comprehension.
The Permanent Methods of Study Involved in the Text.
The Permanent Value of the Content.
The Mechanical Construction of the Text.
The Fundamental Proposition of Textbook Selection.
Data vs. Opinion.
Best Opinion vs. Poor Opinion.
Pooled Judgment.
Neglect Irrelevancies.

CHAPTER II

STUDY ONE: APPLYING THE CRITERION OF INTEREST TO HIGH SCHOOL ENGLISH LITERATURE TEXTS . . . 22

Classification of Books in Order of Interest.
The Readers' Interest is the Criterion.
The Literary Merit Tests of the Lists.
Description of Actual Data.
The Reliability of the Quality and Quantity of the Data.

CHAPTER III

PAGE

STUDY TWO: APPLYING THE CRITERION OF COMPREHENSION TO GEOGRAPHY TEXTS FOR ELEMENTARY SCHOOLS. 66

The Accumulation of the Data.
Selecting Texts to Meet the Needs of Particular Grades.
The Success of Texts in Reaching Their Reading Public.
Individual Differences in Needs of Various Grades.

CHAPTER ONE

PRINCIPLES OF TEXT SELECTION

The major criteria for the selection of text-books are five:

(*a*) The factor of interest.

(*b*) The factor of comprehension.

(*c*) The permanent methods of study involved in the text.

(*d*) The permanent value of the content.

(*e*) The mechanical construction of the text.

These criteria possess determinative importance in the opinion of educators. There is no one way of expressing these criteria. Methods of statement differ but the meat of the matter is constant. We find the above statement convenient for our exposition but claim no particular rhetorical superiority for it.

At present, it is difficult to apply these criteria in a particular situation. The application of general principles to concrete situations has always been troublesome. Educators find themselves in no uncommon impasse when they attempt to apply readily accepted principles to the specific problems of their

profession. The purpose of this monograph is a double one; first to state the fundamental propositions which must underlie sound practice of text-book construction; second to give two researches which are at least moderately successful attempts at such application. The researches are (*a*) the application of the broad doctrine of interest to the problem of selecting texts for high school English Literature work, (*b*) the application of the factor of comprehension to the selection of geography texts for elementary schools.

The history of public education in the United States is very far from a monotonous report of blunders and inadequacies. A study of text-book construction and of the methods of text-book selection even now prevalent, however, gives the student of education reason to prophesy much increase in the genuine adequacy of public school text-books. Further we are sure that the selection of a particular text in preference to other available books is about to become far less a matter of guess and poorly based opinion and far more a matter of deliberate choice in the light of sound criteria of text-book excellence.

Text-book selection is obviously a process of making choices, and choices can be made in many ways. It is quite possible that certain texts have been chosen not because they were better tools of instruction than other books, but because the salesman urging their adoption had a more persuasive vocabulary, more

agile and plastic sales methods, or was politically more canny than his competitors. Other texts have had large sales because they were or professed to be exponents of some pedagogical doctrine which momentarily hypnotized the buying agent. Better than the above, many texts have been adopted by schools because the proper authority, having studied the matter deliberately, chose those particular books.

Fundamental Propositions

The thoughtful school administrator will presumably subscribe to the following four propositions as fundamental to sound practice in text-book selection.

I. Irrelevant differences pertaining to texts should be utterly neglected, viz. pressure from salesmen, exhilarating sales tactics. Texts should be chosen in light of the five determinative criteria of good text construction, namely:

(*a*) The factor of interest.

(*b*) The factor of comprehension.

(*c*) The permanent methods of study involved in the text.

(*d*) The permanent value of the content.

(*e*) The mechanical construction of the text.

II. Objective evidence should be given preference over subjective opinion wherever objective evidence can be reasonably obtained.

III. Wherever subjective opinion is resorted to the final judgment should not be the result of the wisdom of any one person. Final judgment should be the product of the pooled votes of several competent judges. Among these judges should be at least one class room teacher who is to use the book. The more mutually independent the judgments are the more stable the final result.

IV. Wherever subject opinion has to be substituted for objective measurement, especial care should be taken to get the best opinion possible. A very good judge may sometimes give a very hasty opinion.

Objective Evidence

Proposition I, using relevant criteria, and proposition III, using the insight of more than one judge, need no further discussion. The main part of this book is an attempt to apply proposition II to actual, concrete school situations. Objective evidence is usually better than subjective opinion. Many of the deceptions of opinion have been pointed out of late. But the objective evidence has its weak points also. Not all evidence is good evidence. Mere quantity of evidence does not involve in a mysterious way a guarantee that the evidence is reliable or consistent or determinative. Objectivity of evidence never compensates for the errors which come from wrong

methods of gathering the evidence or mistakes in the interpretation of the ever so objective data.

It is hoped that the two problems of text-book selection dealt with in this book will be helpful not only because of the particular findings but also because of the exposition and discussion of the methods used to gain objective evidence concerning texts.

The methods used are of general application. Those who are responsible for text-book selection should know the reading difficulties not only of geographies, but also of all texts, especially arithmetic, histories, and science books.

One of the problems is the determination of the reading difficulty of geographies. Obviously a child with fifth grade reading ability will work to disadvantage if he uses a text which assumes a reading ability rarely found below the seventh grade.

The fourth fundamental proposition of text-book selection is "Where subjective opinion has to be used, care must be taken to obtain the best opinion possible." This means the abandoning of general opinions and the persistent analysis and estimate of the specifics of each text judged.

We no longer rate a school building as a whole. We judge it for ventilation, for play space, for class rooms, for many elements which taken all together make a school house. So with text-books, we should make our judgments not in general, but in particular.

It is apparent that score cards for text-books are useful instruments to aid in making analyzed judgments. Unfortunately such score cards are not in general use. In fact for many texts they are not in existence.

The Use of Rating Score Cards

It is clear that the comparison of one text-book with another will be complete when both texts have been measured in the light of the five criteria. The whole rating of text-books can finally be expressed in mathematical terms on a score card. For purposes of making a decision as to which of several texts to adopt the value of each text can, theoretically, be expressed in one figure. The figure would be the sum of the scores given to each of the five criteria if each criterion were considered as important as every other criterion. Weighted scores would be needed if the criteria were of unequal value. That these five criteria are of unequal value seems to be the case. From a pool of many judgments we assign following point values:

Interest	200
Comprehension	250
Permanent value of subject matter	250
Value of method	200
Mechanical elements	100
Total	1000

Score cards or rating schemes of one kind or another are generally used to get in orderly fashion expert opinion.[1] Many decisions in practical school affairs will have to be based on opinion rather than other data for some time to come. Wherever this is the case it becomes exceedingly important to obtain the best opinion in the best way possible. By best opinion we mean the judgment of those not only who know most about the matter but the opinion of enlightened judges who take the trouble to make the best judgment they can make. Here is where the score card is of value. Opinion expressed through score card ratings will not give insight the judge himself does not possess, but opinion expressed through score card ratings has a better chance of being opinion up to the limit of insight than unanalysed opinion has. Thus we see that the worth of opinion varies in merit not only with the true wisdom of the judge but, also, is a function of the method of obtaining the opinion. Analysed opinion through score card methods increases the chances for a judge to get all his goods to market. The better the score card the better the judgment will be. A Biology Score Card, one of a series, is given here. Its use makes judgment better because it forces the judge to con-

[1] Opinion of any degree of expertness should be resorted to only in the absence of evidential data.

sider many aspects of his problem he might otherwise neglect. An even better score card would make for even better judgments.

Tentative Score Card for Judging Value of Biology Texts

Name of book.....
Author.....
Publisher.....
To be used in.....
Scored by.....

	I	A	I	a
I. Interest	200			
(If there is experimental evidence neglect all sub-topics)				
A. Mention of animals and plants known by student		50		
1. Wild in this district			25	
2. Domestic in this district			25	
B. Possibility of demonstrations which emphasize value of this study to the student		50		
1. In the laboratory			25	
2. Outside			25	
a. Back yard biology				15
b. On excursions				10
C. Style		50		
1. Vividness			25	
2. Clarity			25	
D. Correlation with other subjects		50		
1. Other sciences			30	
2. Literature			10	
3. History			10	
II. Comprehension	250			
(If there is experimental evidence neglect all subtopics)				
A. Reading difficulty		150		
1. Vocabulary			75	
2. Style			75	

TENTATIVE SCORE CARD FOR JUDGING VALUE OF BIOLOGY TEXTS (*Continued*)

	I	A	I	a
B. Difficulty of concepts		50		
1. Degree of novelty			25	
2. Degree of involved presentation			25	
C. Difficulty of graphs and illustrative material		50		
III. Permanent Value of Subject Matter	250			
(If there is experimental evidence neglect all subtopics)				
A. Accuracy and balance of representation of the subsidiary sciences		90		
B. Development of scientific interests		80		
1. Sympathetic but critical attitudes			50	
2. Stimulation to further work			30	
C. Dynamic concept of evolution		80		
1. Material on progress and evolution			40	
a. Development from lower to higher forms in plant and animal life				20
b. Development of nervous system, use of hands and use of voice; and their values to man				10
c. Differences of animals defined as differences of their wants as well as their abilities				10
2. Possibilities for development in the future			40	
IV. Value of method	200			
A. Organisation of material		150		
1. Opportunity to develop general principles			75	

TENTATIVE SCORE CARD FOR JUDGING VALUE OF BIOLOGY TEXTS (*Continued*)

	I	A	I	a
a. Sufficient data to make induction possible				15
b. Summaries				15
c. Indices				15
d. Illustrations and maps				15
e. Graphs				15
2. Opportunity for application of general principles			75	
a. Suggestions for outside work				40
b. Classroom exercises				35
B. Correlation with other subject matter		50		
1. Stimulus to more reading			25	
2. Relation to other work			25	
V. Mechanical elements	100			
(Use objective measurements and norms to judge these values wherever possible)				
A. Size and clearness of print of text		40		
B. Distinctness of pictures		5		
C. Size and clearness of print of foot notes		5		
D. Size and clearness of print of marginal notes and indices		5		
E. Width of margins		5		
F. Length of lines		15		
G. Paper		15		
H. Binding		5		
I. Size and shape of book		5		

CHAPTER II

APPLYING THE CRITERION OF INTEREST TO HIGH SCHOOL ENGLISH LITERATURE TEXTS

Section One

The business of this study is to do two things:

A. *First.*—To name and classify in order of interest the best liked books commonly used in high school literature work. This list will be referred to as the Preferred List.

Name in order of interest the next best liked books commonly used. This list is referred to as the Second Preferred List.

Name in order of interest the third best liked books. This list will be called the Substitute List.

Name in order of violent dislike books commonly used in high schools. This list will be referred to as the Free Reading List.

B. *The second section of our business is to* (1) Defend the proposition that the Preferred List of books should form the basis or core of the English literature work in the high school, that books in the Second Preferred List should be used next frequently, that the Substitute List may be drawn from after the Pre-

ferred and Second Preferred are exhausted and that books in the Free Reading List should never be used.[1]

This may be restated as follows: The best available criterion to use in selecting English literature books is the amount of interest students have in the books, considering only those books which are now in common use.

(2) Present and interpret data calculated to establish the validity of our lists.

The Preferred List

In the second section of this study sufficiently valid data are presented to warrant the assertion that the following are the best liked books commonly used in high schools of Iowa.[1]

Careful students who wish to inquire into the factual basis of our lists before reading the lists will do better if they read the second section first. The busy

[1] Books in the Free Reading List should never be used as a class text. These books should be in the library for the use of those very few students who wish to read them.

[2] Our data, we hasten to point out, are based on generous samplings from high schools of Iowa. Our criterion of interest would apply strictly only to Iowa. The lists of books are generally applicable in so far as any large group of students have about the same interests as large groups of Iowa students. Some exceptions in interest are probably important. For instance *Uncle Tom's Cabin* is highly interesting to Iowa high school pupils. Its interest rating would probably be much lower in Georgia or Louisiana.

school administrator can assume the validity of the evidence and content himself with a perusal of the lists.

TABLE A.—PREFERRED LIST

NAME OF BOOK	INTEREST RATING[1]	RELIABILITY OF RATING[2]
Laddie	97	1.4
Vicar of Wakefield	92	1.7
Kenilworth	79	3.7
Treasure Island	66	2.5
Hamlet	63	2.5
Comus	61	3.5
Uncle Tom's Cabin	61	3.0
Wild Animals I Have Known	60	5.1
Ben Hur	59	3.0
Last Days of Pompeii	59	4.1
The Spy	59	3.6
Romona	58	4.4
King Lear	57	2.2
The Crossing	55	4.3
Snowbound	54	2.5
Luck of Roaring Camp	54	4.0
Macbeth	53	2.2
Ivanhoe	53	3.8
Carlyle's Essay on Burns	53	3.3

[1] Interest rating is in terms of the percentage of students, who, having read the book, declare that it was of distinct

SECOND PREFERRED LIST

Incidents of a French Camp	53	4.2
The Virginian	52	3.8
Les Miserables	52	4.5
The Crisis	51	3.1
The Last of the Mohicans	51	2.6
Up From Slavery	51	5.5
Tom Sawyer	50	2.3
Tale of Two Cities	50	3.0
Silas Marner	49	2.2
Man Without a Country	49	2.4
Huckleberry Finn	49	2.7
Washington's Farewell Address	49	3.3
Ancient Mariner	48	2.2
The Gold Bug	48	2.5
Richard Carvel	48	5.3
Miles Standish	47	3.2
Robinson Crusoe	46	2.6
Lorna Doone	46	4.0

interest to them. Thus the higher the rating, the better the book according to our criterion.

[2] Reliability of Rating is to be read as follows: If many more estimates of the book were taken, the interest rating might go up or down as much as the reliability figure. The chances of its going up or down by an amount greater than the reported unreliability are very few. Thus *The Vicar of Wakefield* has an interest rating of 92 and an unreliability of 1.7, which means that as much more data are obtained, the 92 may go up to 93.7 or down to 90.3. The technique of ascertaining the reliability is fully dealt with in section two.

Free Reading List

In Order of Demerit

Contains books Never to be Used in High School

Paradise Lost	6	1.4
Lay of the Last Minstrel	10	2.5
Heroes and Hero Worship	11	2.8
Essay on Warren Hastings	12	3.4
Macauley's Essay on Johnson	12	2.6
Autocrat of the Breakfast Table	13	2.0
Vision of Sir Launfal	13	1.2
Childe Harold	18	3.5
Webster's Bunker Hill Speech	19	2.8
Fairie Queene	20	2.9
Essays of Elia	21	4.5
Lays of Ancient Rome	21	4.4
Utopia	23	4.5
Thoreau's Walden	23	4.4
Gareth and Lynette	23	2.9
The Princess	24	3.5
Tales of a Wayside Inn	24	2.7
House of Seven Gables	24	2.1
Pilgrim's Progress	25	2.5

Substitute List

Contains in order of probable interest to students books less interesting than the First and Second Preferred List, but of measurably greater interest than the books of the Free Reading List. The books are listed in order of interest merit.

Twelfth Night	46	4.9
Launcelot and Elaine	45	2.6
Hoosier Schoolmaster	45	3.5
Christmas Carol	44	2.5
Hans Brinker	44	3.5
Legend of Sleepy Hollow	43	2.1
David Copperfield	43	2.7
Joan of Arc	43	3.9
Enoch Arden	43	2.8
Lincoln's Gettysburg Address	42	2.2
Leatherstocking Tales	42	3.6
Merchant of Venice	41	2.0
Burke's Conciliation	41	2.9
Midsummer-Night's Dream	40	3.6
Prisoner of Chillon	40	3.1

The interest ratings of 27 other books frequently read in high schools have been determined. The First, Second Preferred and Substitute Lists contain those of highest ratings. These lists contain more books than would ordinarily be used in a four-year high school. It seemed needless to print the others. These interest ratings may be obtained on request.

Notes on the Lists

The criterion used in setting up an order of merit of English literature texts is pupils' interest in the text. This proposition is defended in Section II. Any one whose choice of texts in the practical situation of actual use in high schools is influenced by these lists should bear the following in mind:

(*a*) The interest which students have in the texts is not the only criterion to be used in deciding which one of two possible books to use. Other things being at all equal, however, it is better psychology and sounder common sense to prefer interesting material over material of no interest or positive dislike.

(*b*) Not every book that could be used in high school is included in this study. No book is dealt with here unless it has been reported by at least eighty of the high school graduates who rated the list. Thus we are concerned only with books which are already fairly commonly accepted as worth while. This

matter could be stated: English literature professors and teachers have on the whole approved of many books for high school work. More books are generally approved than can possibly be read by any one student. Of books accepted as useful which shall be chosen? Those of greatest interest to the pupils is our thesis.

(*c*) It is not certain that the lists are equally sound for all sections of the country. In some instances we know radical variations in interest would be very probable. Children's interests must vary with the traditions and interests of the section in which they live; how much is not known. Probably books vary in their relative universality of appeal. Thus Shakespeare may be of about the same interest strength everywhere, whereas *Uncle Tom's Cabin* may have high interest in some sections and practically none in others. We are well within the facts on the other hand to suspect the absence of radical and sharp changes in interest. The children of one county have about the same interests as those in a contiguous county. There is no apparent reason that on a basis of children's interests there should be endless variation within the same state. It might be added that there is probably less difference in interests in city versus rural youths than is ordinarily assumed. To differentiate books on a basis of sex is idle, for boys and girls have to read the same books in most high schools.

There is no reason why our lists might not very closely approximate at least mid-western lists.

(*d*) Other criteria should be used in making up a list for a particular high school. It is sound to hold that other things being equal books in the Preferred List should be read sometime during high school. The books in excess of this Preferred List should be chosen from the Second Preferred List. However, a book with an interest rating of 54% poorly taught will probably do less service than a book with an interest rating of 48% well taught. Thus, while no book should be used from the Free Reading List, in a particular situation should a teacher much prefer a 48% book to a 54% book it is good sense to let her teach it, for the differential of the teacher's attitude is quite probably more important than a small difference in percentage of interest.

(*e*) Should the preponent elements in the situation be a set of slightly lower interest books, on hand, and in good condition, a set of slightly higher interest books available only through purchase, a supply of funds so limited that the purchase of the new books means going without some other good thing; then to delay for a time the desired replacement is probably entirely defensible.

(*f*) The interest which one has in any book is due not only to the character of the book but also to the character of the reader and the method of reading.

Thus the same book though read under equivalent conditions will not have the same amount of interest for all students, not because the book varies but because students differ greatly in what they want out of a book. Differences may lie in many fields such as home training, type of play indulged in, amount of travel, ability to read, type of fantasies or day dreams enjoyed, etc. Thus, even the most interesting books on the list can not be expected to interest *all* pupils. We know they will not. We are pretty sure the books in the Preferred List will very likely be actually disliked by some students. In addition a book which is generally preferred to a second one may in some peculiar instances be less liked by a class or school than the second. However the theory of probability as used in this study is a sound technique in curriculum construction.

Exceptions need not greatly disturb us. Obviously a book in and of itself may be inherently more interesting to pupils than some other. But if the other book happens to be well taught and the book inherently better happens to be grossly mistaught, then the inferior book may very likely appear to be better.

(*g*) A certain variety in texts is undoubtedly desirable. Had the Preferred List on a basis of interest been all drama, or all poetry, or all heroic novel, practical consideration would force us to draw heavily

on other school literature. But the Preferred List as a matter of fact is pretty well balanced as far as kinds of literary content are concerned. And many wide deviations from this list can be comfortably accounted for by the consideration that while what large groups tend to like, other large groups do not tend to dislike, in individual cases no one would expect all pupils to like the same books and in the same relative order of interest. These lists do not pretend to be correct for all students, but as far as we now know they will strongly tend to be better than any other list.

(*h*) It is not useful to think of the most interesting book as the "best" book or the least interesting book as the "worst" book. The most interesting book is *Laddie*. No one would claim that *Laddie* is the best book in the sense that "if only one book was to be read in high school this book should be read." Such choices and such "ifs" do not have to be made. Many books are read. The question is which group is best. In the best group *Laddie* finds a place and *Paradise Lost* does not. This placing is, of course, according to our criterion of interest.

Readers' Interest the Criterion[1]

The criterion used in the construction of these lists is the interest of the student as reported.

No one aware of the administrative exigencies of a high school would contend that "what pupils like" should be the sole criterion for admitting a book, a subject, a technique, or a rule into the total economy of high school procedure. Neither would one suspect that what pupils dislike should be the sole criterion for dropping a book, a subject, a technique, or a rule from the school.

It is hardly conceivable that any teacher is so unenlightened as to hold that disagreeable tasks are worth while because of their disagreeableness. A task may be distasteful and yet worth while, but its worth-while-ness is certainly due to other factors than its unpleasing nature.

Theoretically, most competent English teachers would agree that one of the criteria safely used to select texts would be students' interest, but there would be little agreement on the amount of influence interest should have. If the books available should

[1] Theoretical consideration of the importance of interest is assumed. Personally, we view those who do not appreciate the pragmatic importance of interest in education as in need of instruction in the simple principles of the teaching process.

Nor are we concerned here with the objectives of our English Literature work.

be scored on a basis of 100 points, would literary merit be assigned 50 points, moral teaching 25, and interest 25, or would it be reading difficulty of the composition 50, literary merit 15, interest 15, amount book is used in other schools 20,—or what relative importance would be assigned to interest?

In our present ignorance of the details of scientific management in education, one would utterly despair of getting definite answers to questions of this kind. However answers to such questions must be obtained before we can speak of education as a science in the sense that civil engineering is a science. All would agree that interest is only *one* of several factors which properly influence selection. How influential a factor it should be has not yet been scientifically established.

Our data show that *the selection of English literature in the State of Iowa practically refuses to recognize the factor of interest at all.* The frequency with which books are read in Iowa high schools and the frequency with which students report these books as interesting to them have practically no correlation other than that of chance. This roughly may be taken to mean that the more frequently read books are rated as less than average interest just as often as they are rated as above average interest, also that the less frequently read books receive an interest rating greater than the average as often as "less than the average" rating. Such a state of affairs would be justifiable if it could

be clearly shown that other criteria more important than students' interest determined the frequency with which books are read. That such criteria do not operate will be shown and evidence given later in this study.

Assuming that the 783 reports on which this study is based are a fair sampling, the following examples of the strained relations if not absolute divorce between "how often books are read" and "how much books are liked" will be illuminating.

The 19 books read with the greatest interest were recorded on the average 280 times.

The 20 books read with enough interest to warrant being in a Second Preferred or First Substitute Lists were read on the average 247 times.

The 20 books read with so little interest that they should not be read at all were read on the average 203 times.

The average interest rating of the Preferred List is 62.7%.

The average interest rating of the Free reading List is 16.3%.

Thus, while in frequency the shrinkage in use between the best and worst lists is only 17.6%, the shrinkage in interest is 46.4%! The shrinkage in interest is approximately 2.6 times as great as the shrinkage in frequency of use between the Preferred and the Free Reading List. I doubt if any one would

want to defend this relationship. It certainly would never have happened if the choosers of text-books had known the interest in books before making their choices.

To those who may suspect that a study of only the best books and the worst in point of interest hides a more favorable state of affairs because it does not consider the middle books at all, the relation between frequency of use and amount of interest of all books will be given.

The technique here is that of correlation.

We digress to explain what is meant by correlation. Persons familiar with statistical theory should skip the next page.

The coefficient of correlation is a measure of concomitant variation between two series of facts about the same data. The limits of this coefficient are -1 and $+1$. If the relationship is $+1$ it means that there is a perfect matching. Thus if there were a $+1$ correlation between the frequency with which the books we are studying are read and the amount of interests pupils have in them, then the following conditions would hold:

The book read most frequently would also be the most interesting; the book read least frequently would also be the least interesting; the book read with the fifth greatest frequency would also be the fifth book in interest; the book read next to the least often would

be the book next to the bottom in interest, and so on.

If there were a -1 correlation between frequency and interest then just the reverse would be true, *i.e.*, the most read book would be the least interesting; the book fifth from the top in frequency of the reading would be fifth from the bottom in its interest rating.

If the correlation were zero then there would be no relationship but a chance one between frequency and interest. Some much read books would be high in interest, others low, others average. Some books low in frequency would be low in interest, others high, others average in interest. Thus the sign and size of the correlation coefficient measures the amount of mutual relationship between the frequency with which books in our list are read and the amount of interest pupils have in these same books.[1]

Using only the books which were reported eighty times or more (the unreliability of ratings based on less than eighty gets too large for conservative study) we find the correlation between frequency of use and interest to be $-.04$, practically zero.

[1] Further explanation of the correlation coefficient may be found in

Thorndike. *Mental and Social Measurements.*
Brown and Thomson. *Essentials of Mental Measurement.*
Rugg. *Statistical Method Applied to Education.*
Refer to index in each book.

This means that there is no tendency to read the more interesting books oftener than the non-interesting one. There is also no tendency to read the non-interesting ones any more frequently than the interesting ones.

In view of the fact that the data give also a rating of positive dislike it is instructive to study the correlation between the frequency with which books are read and the positive dislike for them. It should be pointed out that the positive dislike ratings are entirely separate from any interest ratings. They were made independently. The correlation is $-.52$ between frequency of reading and positive dislike. This means that there is a pretty definite factor of negative selection going on at present. The nearest approach to anything like scientific control that the data reveal is this fact that on the whole books which are positively disliked tend to be read less often and there is a slight tendency to read them less often in the order of the strenuousness of dislike.

Certainly the factor of interest should operate in the selection of texts. But it does not. We are not prepared to say that it is theoretically *impossible* to do good work in a subject in which the learner has no interest, but interest is a factor which *facilitates* learning. On the whole, interest in the work runs along with skill in it. With material such as literature the interest factors are probably more important

than in more mechanical subjects where physical skill based on habit predominates. If reading books carries no interest but often even positive annoyance, there is little if any likelihood that reading goes on except under compulsion. The frequent task of reading inherently uninteresting material under school compulsion may account for whatever truth there is in the general feeling that adults on the whole read little literature of great merit. Very few high school graduates ever read the *Æneid* over again for the fun of it. We doubt if one high school graduate in fifty, who, having read *Sesame and Lilies* or *Paradise Lost* in high school, ever reads these books again, or reads other works by the same authors unless they are in turn teaching them in high school for a wage, to pupils who will never look at them or think of them again unless later as high school teachers, they in turn—and so on.[1]

Several cross sections of the data are presented here to further illustrate the divorce between interest and frequency of reading certain books.

Table B contains the 29 books read the most frequently. On these we have 350 reports or more.

[1] The way a book is taught must determine in part the interest pupils have in it. This relation between how a book is taught and resulting interest is as yet unknown.

Table B

No. of Times Reported	Name of Book	Place on Interest Rating[1]
720	Vision of Sir Launfal	Free Reading List
607	Merchant of Venice	Substitute
552	Legend of Sleepy Hollow	Substitute
535	Lady of the Lake	No place
529	Silas Marner	Second Preferred
523	Macbeth	First Preferred
523	Lincoln's Gettysburg Speech	Substitute
503	Ancient Mariner	Second Preferred
492	Ivanhoe	First Preferred
481	Courtship of Miles Standish	Second Preferred
465	Great Stone Face	No place
460	The Raven	No place
459	Julius Caesar	No place
424	Man Without a Country	Second Preferred
414	Gold Bug	Second Preferred
411	House of Seven Gables	Free Reading List
397	Christmas Carol	Substitute
385	Last of the Mohicans	Second Preferred
383	Canterbury Tales	No place
383	Snowbound	First Preferred
380	Passing of Arthur	No place
378	Sketch Book	No place
375	Hamlet	First Preferred
369	Treasure Island	First Preferred
368	Thanatopsis	No place
368	Idylls of the King	No place
366	Robinson Crusoe	Second Preferred
366	Gray's Elegy	No place
357	Launcelot and Elaine	Substitute

[1] "No place" means that the book has an interest rating better than the Free Reading List but not high enough to reach even our Substitute List.

Of the 29 most frequently read books (350 reports or more out of 783 students)

5 are in our First Preferred List
7 are in our Second Preferred List
5 are in our Substitute List
10 are in "Not Placed" list
2 are in the Free Reading List

Total 29

Of the ten books which had an interest rating of 45, one was read 640 times and one but 160 times. The other eight were rather evenly distributed between these extremes. Of the twelve books read with an equal frequency (280 times) one had an interest rating of 10, another of 95. And here the other ten were scattered between these extremes.

All the data show how books with equal interest are read with highly variable frequency. Books read with equal frequency have highly variable interest value.

So far we have shown that interest as such exerts no measurable influence on frequency of reading over the state of Iowa as a whole. We wish to present data to show that the operation of other definite criteria is not evident.

If clean cut criteria were operating under conscious direction some one would know about it. Who does? What are the criteria? Reading difficulty is a very good criterion, but the reading difficulty of books

used in high school English literature has never been determined. Some one could very well determine this.

Perhaps information value is thought by some to be an effective criterion which we have carelessly overlooked. We have yet to see a study of the information value of the books under consideration. Such a study might well be made. At present, however, this criterion is still a "poor relative" of opinion and a very much be-fussed relative into the bargain. It would be difficult to show that the much read books are higher in information value than the much liked books. Comparing the First Preferred List with the Free Reading List one might perhaps be impressed with the realism of the Preferred List in contrast to the fantasy of the Free Reading List. Possibly people have to be older than of high school age before they become sufficiently disillusioned about the world to find pleasure in fantasy. The high school pupil may be still satisfied with the world as it is, or nearly is. He may not yet be interested in Utopias, Paradises Lost or Found, Fairie Queens or Visions, or he may be too old to be interested in them.

LITERARY MERIT NO CRITERION

It is always possible to fall back on the well known defense of literary merit. It could be asserted, if one is careless enough about the facts, that the criterion of literary merit decides the frequency with which books on the whole are read in high school.

In one sense this is true. Probably no book is much read until by a vague consensus of opinion it is worth reading. But this does not effect our data because we have considered only books rather frequently read. They all possess enough literary merit to be generally used. But differences in literary merit in excess of a minimum amount of it to allow entrance of books into common use can not account for the frequencies.

We have accepted literary merit in this sense, as a working criterion of a minimum amount. It is important to realize that in no other sense can we hold it to be influential. It is not at present a criterion for two reasons. First, there is no objective measure to use as a criterion. There may be some day a scale for rating for literary merit. Up to now contributions on the matter have been much cluttered with opinions and not disciplined with scientific or orderly study. In short, every one has a literary rating all his own, (which rating we suspect is largely a rationalization of likes and dislikes by the individual doing the rating).

Second.—If differences in literary merit decided differences of frequencies of use so that the most meritorious book was read the most frequently, the second most meritorious book was read the next most frequently, and so on, then there would have to be a *general agreement* as to the differences in literary merit of the books concerned. There is not even an approximation of agreement, as the following data show:

We have had five groups of judges rate the nineteen books composing the Preferred List and the nineteen books composing the Free Reading List for the quality, literary merit. The judges were in all instances ignorant of the interest value of these books. They were urged to give their own opinion independent (as far as possible) of fame of the author, popular appeal of the book, etc.

The value A was to be given to those books which were of preeminent literary value, the great books of the ages.

The value E was to be given to those books which were of such poor quality that probably they should not be considered for serious work at all.

The value C was to be given to books about half way between A and E.

The value B was to be assigned to books better than C but of less merit than A.

The value D was to be assigned to books worse than C but better than E.

The additional instruction was given that in case of unfamiliarity with a book no rating need be given.

The pooled results of these judgments of the literary value of the Preferred List (each book given a number) and of the Free Reading List (each book given a number) is presented. The reader is asked to distinguish on a basis of literary merit which list of books is the best.

List X[1]

Showing the distribution of judgments on nineteen books for literary merit. A, B, C, D, E, amounts of merit as defined in text. Numbers 1, 2, 3, etc. refer to books.

	A	B	C	D	E	T
1.	3	21	27	6	3	60
2.	9	41	30	6	1	87
3.	21	27	7	2	0	57
4.	16	56	43	4	0	119
5.	52	46	16	11	2	127
6.	3	34	22	10	1	70
7.	25	30	38	22	2	117
8.	24	32	10	15	10	91
9.	51	31	36	2	2	122
10.	14	28	49	6	7	104
11.	9	33	38	13	3	96
12.	22	26	39	13	4	104
13.	48	31	26	5	1	111
14.	14	22	21	20	4	81
15.	30	33	47	7	1	118
16.	9	31	20	12	3	75
17.	53	37	16	6	1	113
18.	18	51	29	4	1	103
19.	7	17	21	12	6	63
T	428	627	535	175	52	1817

[1] Whether this list is the Preferred or the Free Reading List will be stated later on. This postponement is to enable the reader to see the utter hopelessness of getting a "literary merit" when the books are given by number rather than name.

List Y

	A	B	C	D	E	T
1.	27	44	18	1	3	93
2.	3	19	27	2	0	51
3.	25	29	24	4	3	85
4.	6	2	9	16	11	44
5.	9	36	33	4	3	85
6.	51	25	18	3	0	97
7.	21	52	20	3	1	97
8.	3	27	31	7	4	72
9.	3	31	35	4	0	73
10.	9	29	23	7	3	71
11.	8	6	19	3	3	39
12.	10	22	8	4	3	47
13.	36	21	13	14	3	87
14.	18	31	15	3	1	68
15.	1	25	23	2	2	53
16.	15	35	24	2	3	79
17.	39	26	35	5	1	106
18.	12	17	44	7	6	86
19.	30	37	31	5	2	105
T	326	514	450	96	52	1438

The reader will readily see that a great amount of sagacity would be necessary to determine whether List X or List Y is superior in literary merit. There is certainly not enough difference to outweigh the essential difference in interest value which exists between the two lists.

The careful student will immediately wonder if our pooled ratings are not a mixture of inexpert and consequently valueless opinion. We shall therefore spread out the pith of each group of ratings to show that there is no difference of importance between the judgments of groups which vary in expertness.

Group I is a class in psychology of advertising. This group is not expert in matters of literary merit of English literature. Their judgments are:

List X of 369 judgments on 19 texts	List Y of 230 judgments on 19 texts
21.4% judgments were A	12.2% were A
35.6% judgments were B	38.7% were B
31.5% judgments were C	42.2% were C
9.5% judgments were D	4.3% were D
1.9% judgments were E	2.6% were E

We interpret these data as showing that in the opinion of a psychology of advertising class List X is slightly superior to List Y in literary merit.

Group B is a class of college juniors preparing to teach in high schools. We assume that their judgments are slightly more expert than the judgments of Group A.

List X	List Y
of 995 judgments on 19 texts	of 803 judgments on 19 texts
28.3% of judgments were A	19.9% were A
36.8% of judgments were B	39.9% were B
24.7% of judgments were C	29.0% were C
6.6% of judgments were D	8.5% were D
3.5% of judgments were E	3.6% were E

In the judgment of Group B List X appears to be slightly better than List Y. The chief characteristic of this difference is however its negligible smallness. Other criteria than literary merit should obviously operate in deciding which list to use.

Group C is a graduate seminar in educational psychology. Most of the group are experienced teachers. They are presumably still more expert than either Group A or B.

List X	List Y
of 216 judgments on 19 texts	of 222 judgments on 19 texts
26.3% of judgments were A	36.5% were A
35.6% of judgments were B	31.1% were B
27.3% of judgments were C	23.9% were C
9.3% of judgments were D	6.3% were D
1.4% of judgments were E	2.3% were E

Here List X has 61.9% of judgments B and A
List Y has 67.6% of judgments B and A.

There is a slight difference in total literary merit in favor of List Y. But the difference is so slight that no one could possibly judge it to be as important psychologically, pedagogically, or culturally from the teaching standpoint as is the difference in interest value determined by this study.

Group D, our most expert group, is Doctor Craig's seminar in introduction to graduate work in English. Many of the group are instructors or have been instructors in English literature. They have profited by unquestioned expertness in English literature through graduate study in English.

List X	List Y
of 182 judgments on 19 texts	of 176 judgments on 19 texts
22.5% of judgments were A	14.2% were A
21.9% of judgments were B	27.8% were B
34.6% of judgments were C	38.6% were C
15.9% of judgments were D	14.7% were D
4.8% of judgments were E	4.5% were E

Two facts are apparent: (1) the expert group rated both lists on the whole lower than the relatively less expert groups rated them; (2) in the judgment of the most expert group List X and List Y are practically of equal literary merit. Perhaps List X is slightly better. *In all cases List X is our Preferred List and Y our Free Reading List.*

The force of the argument is now apparently adequate.

When groups of judges varying in expertness are asked to pass on two lists of books the experts hand in about the same findings as the less expert. There is then no expert opinion as different from common knowledge. For if there were a *useful* expertness graduate students in literature would have it. They do not. Our blunt conclusion is: since judgments of literary merit on our Preferred List as a whole and on our Free Reading List as a whole can not be made to even *suggest* significant differences, then literary

merit as a criterion, except as we have already admitted it[1] is not determinative. This we take it leaves a clear field for the validity and usefulness of the lists presented in the first of this study which are based on the criterion of interest.[2]

This lack of agreement in matters of merit is not surprising. After all literary merit or any of its synonyms contains but little useful truth. It is similar to such phrases as the Well-rounded Life, Education for Complete Living, True Culture, Good Citizenship, and the like. We assume that any phrase such as those contains a notion upon which many can agree as they can agree on the notion that two plus two equals four. No such notion is contained by such a phrase as Literary Merit, or Education for Complete Living, or Good Citizenship. These are verbal fog screens concealing chiefly differences of opinion. After all what may appear to you as a well-rounded life

[1] Minimum amount to be admitted at all.

[2] If the reader should feel that there are a few books in the Free Reading List of such great merit that they should be read in spite of their low interest value he can satisfy himself as to the difference of opinion on the matter by attempting to pick out from the pooled judgments of List Y *which books they are* by the spread of judgments presented there. He might possibly pick number 6, 13, 17, which are *Autocrat of the Breakfast Table, Utopia, Tales of Wayside Inn.* No one of these books are really liked by one pupil out of four who have read them.

may appear to your neighbor as the life of a prig, a selfish hoarder, a silly person, or what not. While his notion of a well-rounded life may seem to you to be a life too physical, too intellectual, too ethereal, or too worldly. And all the while we know you are both wrong and that our notion of the well-rounded life though sharply differing from both of yours is really correct.

So with literary merit, perhaps that is meritorious which is liked or which satisfies, or in some other way leaves a characteristic impression on the reader. As you and your neighbor are forever satisfied with different things, literary merit is one thing to you and another to him.

Even though this explanation is faulty the cold fact remains that there is as yet too little agreement as to the varying worth of the books we are considering to use merit as a criterion.

Third,—if this were at present a criterion, it certainly is not effective. For who would say the following differences in frequency of reading were due to clear differences in literary value:

Vision of Sir
Launfal.....read 720 times out of 783 reports
King Lear.....read 185 times out of 783 reports
Merchant of Venice.read 607 times out of 783 reports
Hamlet.....read 375 times out of 783 reports
Julius Caesar.....read 459 times out of 783 reports
Treasure Island....read 369 times out of 783 reports
Sesame and Lilies...read 145 times out of 783 reports
Gold Bug.....read 414 times out of 783 reports

Another defense for the present situation would be that a balanced literary ration must be provided for high school pupils and the present practices are approximations of this balance. The necessity for balance is of course entirely proper and sound. But there is some difficulty in justifying the wide spread notions about just what a balanced ration is. However, it is to be noted that there is at least variety in the Preferred List as well as in the other lists. There is a fairly good variety both as to the type of literature, date of publications (*i.e.*, age) and many of the books in our lists have undisputed literary merit. We doubt if in the matter of balance the books most often read make any better presentation than do the books which when read are of distinct interest.

The present situation in English literature as far as the frequency with which books are read is concerned, is certainly not the result of conscious, purposeful, scientific control and direction, right or

wrong. It is a result of a confusion of influence hopelessly interwoven. In addition it should be kept clearly in mind that even if other criteria were operating still the present impotence of the interest factor can be viewed with philosophic calm only by the pedant.[1]

The evidence so far supports the following statements:

The correlation between the criterion of interest and the frequency of reading is zero.

The criterion of reading difficulty can not operate in a useful fashion for the difference in reading difficulty has never been determined.

The criterion of useful information is not operative for the useful information rating of the books is too nebulous.

Literary merit doubtless operates in the restricted sense of determining in some way the entrance of a book into use at all. But it does not determine the relative importance of books that are used, for there

[1] It can be shown statistically that the operation of definite criteria need not be assumed to account for the present situation. For the frequency with which books are read conform to the upper half of a pure chance curve within the probable error. The treatment of the data to prove this is deleted for we do not need it to prove our point concerning the absence of steering and the prevalence of drift at present. Moreover our interest is not to prove things are wrong but to sell better English literature choices to school managers.

is not sufficient agreement as to the measuring of literary merit.

We repeat that the use of some criterion though imperfect is better than the vague and uncertain practices which now exist. If choices of high school literature texts were made in the light of the lists of this study the situation would not be made worse and would probably be genuinely improved. The criterion of interest is not only known but is of respectable importance in any teaching situation.

Section II

We have presented lists of literature texts.

We have defended the use of the criterion of interest in setting up lists.

However, the lists are good only to the extent that the data upon which they are based are valid. We now present a description of the original data.

The freshmen English classes at the State University of Iowa were given as an exercise the filling out of the following blank.[1]

We took freshman data not because those only were available, but because freshmen appeared to us to be the proper judges. High school seniors are a little too near some books to

[1] The data were obtained through the English Department of the State University of Iowa of which Professor H. C. Craig is the head. For this courtesy we desire to express our appreciation.

have a good perspective. College seniors' ratings would have been the result of original preference plus four years of college training which would certainly tend to change their attitude toward some books. Seniors would tend to rate not on what they did like but on what they now think they ought to like. There would be a repetition of the "old oaken bucket" delusion. The necessity of carrying water for mother, the weight of well sweep, the leaks in the bucket, the spilled water, the ice on the well, all tend to be forgotten as distance softens the hard spots and enchants the pleasantness of childhood. I suspect that many honest grown-ups *really* think that in their youth they said at story time "Mother, tell us a story and be sure to put in a wholesome moral."

QUESTIONNAIRE

1. Check in the first column those books which you studied in class during your high school course.
2. Check in the second column those books which you read out of class as collateral reading.
3. Check in the third column those books which you now believe to have been of the most interest.
4. Check in the fourth column those books which you now believe to have been of no particular interest, and yet not entirely lacking of interest.
5. Check in the fifth column those books which you now believe to have been of no interest whatever, and worthless.
6. Check in the sixth column those books which you would drop from a course.
7. At the bottom of the last page you will find space. List here a number of books you would add to a course.

Name.....
High school last attended.....
Major study in S. U. I.....
Age.....

1	2	3		4	5	6
			Burke's Conciliation of the Colonies.			
			McCaulay's Essay on Ben Johnson.			
			Macbeth.			
			King Lear.			
			Midsummer-Night's Dream.			
			Twelfth Night.			
			Hamlet.			
			The Ancient Mariner.			
			Silas Marner.			
			Ivanhoe.			
			Vicar of Wakefield.			
			Snowbound.			
			Roger De Coverly Papers.			
			Irving's Sketch Book.			
			Great Stone Face.			
			The Last of the Mohicans.			
			The Vision of Sir Launfal.			
			The Courtship of Miles Standish.			

etc.

This list contains 108 books.

Returns were received from 732 students. These data were tabulated in ways to permit of convenient study.

The correlations which are reported in this study were computed by the Pearson formula

$$r = \frac{\Sigma xy^1}{N\sigma_x\sigma_y}$$

There are, of course, certain objections which can be raised about the trustworthiness of the data. I shall consider first the trustworthiness of the data as to the quantity, then as to the quality of it.

Quantity of the Data

The ratings of the books in this study can be accepted with confidence as far as the amount of evidence is concerned. We shall discuss at some length the reliability of the ratings because they are expressed in percentages. The reliability of percentages has so far been too little considered in scientific studies in education. 732 questionnaires were returned. Not all the books were read by every one of the 732 students. In fact, no book was reported on by all. Obviously the more cases there are in any study the greater the reliability of the findings, other things

[1] See Thorndike *Mental and Social Measurements*, page 173.

being at all equal. We could have wished to base our interest ratings on 7320 cases instead of 732. Our evidence is much stronger because it is based on 732 reports than it would have been if only 73 students had rated the books.

Since the reliability of the ratings is a function of the number of reports studied we want to know just how reliable our findings are. To illustrate: *Hamlet* was judged by 375 students. 63% judged the book to be of genuine interest to them. Can we generalize from this and say 63% of high school pupils taken at large find *Hamlet* of distinct appeal to them? Of course the ideal way to determine what percentage of high school pupils find *Hamlet* of especial fascination would be to ask all of those who had read *Hamlet.*

Less ideal than this way but far better than guessing is the determination of the reliability of the 63%. This can be done by the application of the formula of simple sampling. Each one of the 375 students, who having read *Hamlet,* rated it, can be thought of for our purposes as performing a little experiment. He is to rate *Hamlet* as interesting or not interesting. 375 rate *Hamlet* and 63% of them rate it as interesting. Suppose we went on and got 700 ratings. Would 63% of the 700 rate as did the first 375? The percentage of 63 would probably vary slightly. But variations would probably center around 63. After

700 or 1,700 more students had been asked we should expect their percentage of interest to be nearer 63 than 12 or 15 or even 50 or 90. That is, on our 375 cases we should expect that the interest ratings of a much larger group would not be exactly 63 but that 63 would be the best single bet we could make in the light of our facts.

The formula $\sqrt{npq}$ when n = number of cases, p = % of pupils who rate *Hamlet* as interesting, and q the % who rated the book as not interesting, will tell us the standard measure of variation from 63 which the percentage of interest will take as more cases are added.

To illustrate the process in detail:

n = 375 students rated *Hamlet.*
p = 63% of these rated *Hamlet* as interesting.
q = 37% of these did not rate *Hamlet* as interesting.

With 375 cases to predict from how will the percentage of interest probably vary as many more cases are added?

This is found by applying formula $\frac{\sqrt{npq}}{n}$.[1]

In our illustration $\frac{\sqrt{npq}}{n} = 2.55\%$.

[1] The quantity $\sqrt{npq}$ is itself divided by n to change it from a quantity in terms of n to a percentage.

This means that as more cases are added 63% may go up 2.55% or down 2.55%. The probability that the % of interest will not go over 65.55% (63 + 2.55) or below 61.45% (63 − 2.55) are 68 chances out of 100. The probability of the interest in *Hamlet* being less than 57 or more than 69 is but two chances out of 100. It is evident that, when only those cases where n is 80 or more are involved, the reliability of our findings is very high indeed. So much for the validity of our data as to quantity.[1]

It is possible that we have enough data but its quality is such that it is untrustworthy. Here the argument would run as follows. If you ask any number of people a question about which they know nothing the answers are no better than guesses. The average of many fake answers is not a right answer except by chance.

College freshmen may not know why they liked and disliked certain books in high school, but it is reasonable to assume the plain facts that pupils do like and dislike books and that they know what interests them and what does not. If their reports are false they are false not because knowledge was

[1] Those who wish to further satisfy themselves on either the soundness of the theory of "Sampling" or on the technique involved will find Yule's *Introduction to the Theory of Statistics*, Chapter XIII, a satisfactory reference.

lacking but because they purposely lied or reported with hopeless carelessness.

There was certainly no reason for deception. There may have been a slight tendency to say the proper thing which in this case would be to report more favorably than their true interests warrant on books which are famous, or old, or "high brow" and to underrate their true interest in the so-called popular books. The fact that the percentages of positively disliked books are relatively small may perhaps be interpreted as a result of two factors:—genuine dislike minus a hesitation of going on record as disliking books that one "ought" to like. To the extent that our data are vitiated by this hesitation, it gives a case weaker not stronger than the true case in favor of our Preferred List and against the Free Reading List.

The following relationships of the data are our evidence of the validity of the data.

The correlation between interest and a neutral attitude toward the books was — .18.

The correlation between neutral attitude and dislike was +.51.

The correlation between dislikes and dislike so great that dropping the book in question was recommended was +.67.

These three correlations go right against any notion that the ratings were the product of carelessness. In

fact, if the markings were the product of carelessness rather than of thought we simply could not account for these correlations.

The correlation between interest and a neutral attitude (−.18) is practically no relationship at all other than a chance one. And it is about what one could reasonably expect to be the fact. The great interest that some have in a book can very well run along with a neutral, 50-50, attitude of others. Unless one is under the spell of the "pathetic fallacy" he would not expect to be able to establish any high correlation between the indifference of one group and the interests of another.

He would however expect more alienation to exist between the interests of some high school pupils and the positive dislikes of others than between the interest of some high school pupils and the neutral attitude of others. There is nothing in the nature of children or in education which creates a tendency of likes of some to correlate with the dislikes of others more closely than with a neutral attitude. Just the opposite would be more likely. This is exactly what we get.

There is a negative correlation of significant size between likes and dislikes (−.377); there is none between likes and neutral attitudes. This inner consistency of the data argues potently for the validity of the data. If likes and dislikes had correlated posi-

tively instead of negatively it would cast a severe criticism against the quality of the data. They correlate negatively.

The correlation between neutral attitudes and dislikes was +.510. This shows that there is some positive relationship between these two series of ratings. What some rate as of indifferent value, others tend to rate as disliked. A common sense interpretation of this might be that in expressing opinions, indifference of some predicts dislike of others. Perhaps timid souls mark as indifferent what more valiant spirits rate as disliked.[1]

One other correlation further leads us to maintain that our data were honestly given and represent the facts—the correlation between dislike and great dislike, *i.e.*, enough to advise dropping from the course, is +.67. This can only be interpreted if we accept the validity of our data. Surely there ought to be a high positive correlation between a tendency to dislike and a tendency to dislike a great deal. This relationship ought to be greatest of any in the data. It is the greatest relationship.

Every correlation we have fits in with a notion that the pupils gave a fairly true report of their attitude

[1] If one wished to make a general observation of human nature concerning the reports of people it would be: indifference is more closely related to dislike than to positive like or interest as we use the term.

about books read. *If their ratings had been false or overcareless no such consistency of relationship would have been at all probable.*

The data upon which the lists are constructed have a good reliability for two reasons.

(*a*) The quantity of the data is too great to admit of a distorted picture due to inadequate sampling.

(*b*) The inner consistency of the relationship between

Interest and a neutral attitude	(= −.184)
Interest and dislike	(= −.377)
Neutral attitude and dislike	(= +.510)
Dislike and great dislike	(= +.670)

gives us plenty of evidence to accept the qualitative superiority of the data. All the original data and subsequent computations used in this study are on file in the College of Education, State University of Iowa, Iowa City, Iowa.

CHAPTER III

APPLYING THE CRITERION OF COMPREHENSION TO GEOGRAPHY TEXTS FOR ELEMENTARY SCHOOLS

Comprehension Values of Geography Texts

All the tables used in this study are put together at the end. It is believed that this will facilitate the study of this research.

This report, as part of the judgment of relative value of geographies, deals with the comparative difficulty which children find in reading the various texts understandingly. The method used in this study is of general use. No fundamental change in it is necessary were we dealing with history or science or other texts. We would define the comprehension value of a text as the degree to which children may read passages of that text and answer simple questions on such excerpts without entering into any intricate consideration of the difficulties involved in comprehending the reading matter. We can then arrive at a very adequate notion of the quality which we are comparing. With such a formulation in mind, we have formed random samplings of the reading matter of each text book into tests and have asked simple

questions on the basis of each sampling. We will then judge the value of any one book by children's responses to questions on its subject matter.

There were five series of books which were candidates for selection. The series were lettered and each first book was called "1," and each second book was called "2." The names of the five series are deleted for obvious reasons. Throughout this study they will be referred to as series A, B, C, D, and E.

One could not hold that reading difficulty or ease should alone decide which text to use. The other criteria of selection must be given due weight. Books of reading difficulty too far above or below the reading ability of the pupils who use the books are undesirable. Quite often superior subject matter necessarily involves slightly greater reading difficulty. For instance, a superficial description of the "tides" may involve easier reading than an adequate discussion of the ocean movements. Should the more difficult book be used the teacher should know where the hard passages are and govern herself accordingly.

In order to ascribe to the book itself the adequacy or inadequacy of the responses of children to questions asked, after the reading of any portions selected from it, we need to be perfectly sure of three major conditions that may influence the behavior, aside from the actual difficulty of the material in the text. We must know:

1. That the samplings are a fair representation of the book and that the questions on the passages

from one book are no more difficult than the questions on passages form any other book.

2. That the children who answer questions on one book are just as capable as those who answer questions on any other book.

3. That the order in which the tests are given has no effect on the responses of the children who take them.

The data which follow are organized to present results which adequately account for these three factors, and consequently we may attribute the values, as indicated through the responses of the children directly to the texts.

The Fairness of the Sampling

1. If the responses of the children to the questions on any one book were materially affected by chance inclusions of difficult passages, there would be no relation between the scores made by a group on the first halves of the tests and the scores made by the same group on the second halves of the tests. A group might then get a high percentage of the questions forming the first half of the selections correct, and a low percentage on the second half of those selections correct. On the other hand, if on the whole the books which are acclaimed as easy by selections forming the first half of the test (which are taken from the first

half of the text) are also acclaimed as easy by selections forming the second half of the test (which are taken from the second half of the text), then we can safely decide that the difficulty of the selections is a function of the book. Also such agreement would indicate that the general difficulty of the questions has not been greatly influenced by the mood or ability of the author of the test. Agreement between the average difficulty of the first half with the average difficulty of the second half of the test, shows that the author's personality was not a determining factor in this difficulty.

It could be maintained that all of the questions to one book were easy, whereas all of the questions to another book were hard, but we do not believe that this is a serious possibility because of the method used in construction of the tests. The tests were, of course, constructed with strict impartiality—passages being taken from the same pages of each book, and crucial questions being asked from each passage, with no consideration except the logic of the passage, as the author understood it.

Table 1 gives the agreement as it is. It will be noted that the books which are easy to read, by the verdict of children's responses to the first halves are also easy to read by such verdict of the second halves. Those that are hard to read when we consider the first half are hard to read when we consider the last half.

There is only one exception, and that is D-1 and D-2. The first halves of these two books are easy to read, and the second halves are hard to read. It is important to note that these two exceptions are of the same series. Those books were written in that way.

The correlation between values by the first half and values by the second half is +.32. The correlation of values by the first half with values by the total is +.76. The correlation of values by the second half with values by the total is +.83. Though correlations with cases as few as ten are unallowable, when the object is to show relationship between qualities or between tests, it is a valuable means of showing characteristics of actual data, no matter what the number of cases. It would be unjustifiable to say that first halves of tests such as these correlated as a rule with second halves about .32, but it is entirely justifiable to say that these actual first halves do actually correlate .32 with these actual second halves. By Brown's formula between the totals with totals obtained in the same way it goes up to .48.

It will be profitable to note in this connection that each second book of a series is harder than its first, and that the volumes I place the series about as the volumes II do, except the E-1 and E-2, and that of all the series, it would be most likely that the difficulty of the first volume of the E series had a different relation to the other first volumes than the second volume of

this series had to the other second volumes, since Mr. X. wrote the first volume and Mr. Y. wrote the second, whereas in all of the other series, volumes I and volumes II were both written by one or by both men.

The Equality of the Groups Used in the Tests

2. Children whose responses are being evaluated to gain comparative estimates of the values of these texts are all twelve-year-olds. Since they are all of the same age and since there are from about 350 to about 500 in each group, the abilities of the groups must be about equal, provided that their verbal intelligence is about equal. Table 2 gives the distributions on Wylie Opposites test A, B, and C, given to these same children about one year before the Geography Reading tests were given. These children were then eleven years old, and Table 2 represents the distributions of Wylie scores for eleven-year-olds of the same schools that housed the twelve-year-olds who took the geography tests. "A" represents distributions of Wylie scores of eleven-year-olds January 14, 1921, in the same schools where the twelve-year-olds took the Geography Reading tests, A-1 and A-2, early in 1922. "B" represents distributions of Wylie scores for eleven-year-olds January 14, 1921, in the schools

where the twelve-year-olds took the Geography Reading tests, B-1 and B-2, early in 1922, etc. Table 3 gives these same facts for twelve-year-olds, January 14, 1921, by groups of schools which took the various Geography Reading tests. Table 3 gives the facts which insure us that it is not characteristic of any of these groups of schools to have twelve-year-olds which are much brighter than eleven-year-olds, and that therefore the relative abilities of the twelve-year-olds in these groups of schools will be just about what those same children were one year before, when they were eleven-year-olds.

In Table 2 the percentage of zeros was for the "A" group, .050; for the "B" group, .070; for the "C" group, .070; for the "D" group, .067; for the "E" group, .058. The lines punctuating the distributions are percentile lines. The first line in each distribution marks off the lowest 10% from the other 90%. The next line marks off the lowest 25% from the highest 75%. The third line marks the lowest 50% from the highest 50%. The fourth line marks the lowest 75% from the highest 25%. The last line marks the lowest 90% from the highest 10%. The remarkable correspondence of the position of these lines in the five groups in both Table 2 and Table 3 shows beyond any doubt that the groups of children are identical, so far as ability on the Wylie test is concerned.

Negligible Effect of the Order of Giving the Tests

3. The possibility that the order in which tests were given might affect the responses of the children and would therefore introduce a factor in the evaluation which could not be justly attributed to the book itself, needs consideration. It has been repeatedly found that tests constructed as these were, are not easily amenable to practice. We found at Garden City that giving practice in this function so as to make an appreciable difference in the average of the groups took two years of special training; and Dr. Arthur Gates found at Scarborough, where the abilities were already near the intelligence limit, that practice in the function made no difference at all in the test.

All the "2's" were given directly after the "1's," and still each "2" is harder than its corresponding "1." This seems to indicate that the order of giving has no great influence on the results. A-1, B-1 and C-1 were given before any other test, and D-1 and E-1 were given after one of the other three, both volumes of the series, had already been administered. In spite of this fact, B-1 and C-1 are pronounced the two easiest, and A-1, the hardest to read. This would seem to indicate that the practice gained from previous experience in the test had no great influence on the results.

Some schools were chosen in which to change the order of the giving of the tests. In considering these results we must bear in mind that the determination for a single school[1] is very unreliable, and if the general tendency of the results is like the results of the total, in spite of the change in order of administration, our conclusions will be verified. In No. 1 School the order of giving was E-1, D-1, C-1, B-1,—with the following results:

Text	Per Cent of Correct Responses
B-1	.76
E-1	.67
C-1	.66
D-1	.63

There were seventeen 12-year-olds.

In School No. 2 the order of giving was E-1, E-2, D-2, C-2,—with the following results:

Text	Per Cent of Correct Responses
E-2	.75
E-1	.70
D-2	.69
C-2	.57

There were one hundred and fourteen 12-year-olds.

[1] The data used here were gathered in a middle western city.

In School No. 3 the order of giving was D-1, E-1, C-2, A-2, with the following results:

Text	Per Cent of Correct Responses
E-1	.81
D-1	.75
A-2	.67
C-2	.66

There were forty-six 12-year-olds.

In School No. 4, the order of giving was D-1, D-2, E-1, C-2,—with the following results:

Text	Per Cent of Correct Responses
E-1	.67
D-1	.65
D-2	.57
C-2	.54

There were sixty-eight 12-year-olds.

In School No. 5, the order of giving was E-1, E-2, C-2, D-2, with the following results:

Text	Per Cent of Correct Responses
E-1	.59
D-2	.58
E-2	.57
C-2	.51

There were fifteen 12-year-olds.

In School No. 6, the order of giving was D-2, E-1, E-2, C-2,with the following results:

Text	Per Cent of Correct Responses
E-2	.39
E-1	.37
C-2	.27
D-2	.22

There were nine 12-year-olds.

Certain noticeable features deny "the order of giving" any part in the results. Certainly the test which is given last is not always the easiest. The only case in which it is easiest is in School No. 1, and there it is because B-1 was given last, and B-1 is by our total result the easiest. Also there is no case in which the test which is given first is demonstrated to be hardest. E-1 is always about as hard as D-1, as it is in our totals, no matter what the order of giving. As a matter of fact, E-1 was easier for the children of School No. 3 School No. 4 and School No. 5 than D-1, even though E-1 was given first in No. 4 and No. 5, and D-1 was given first in School No. 3. A-2 bears the position in difficulty in School No. 3 that it has in our totals, even though it was given last. C-2 is the hardest for the children at No. 1, No. 3, No. 4 and No. 5, and next to the hardest for the children of No. 6, even though it was given last, or next to the last in each,—that is to

say, C-2 which is hardest by our summaries when practice in only one test (C-1) had been given, is also hardest in these special schools when practice in two or three tests had been given before the administraction of C-2. E-2 is easier than C-2 and D-2 in No. 2 and No. 6 even though it is given before C-2 and D-2, and is about as hard as D-2 and easier than C-2 in School No. 5, even though it is given before C-2 and D-2.

This analysis of the schools where a special order of giving was instituted seems to indicate that the data are independent of the order in which the tests were given.

How Reading Difficulty of the Texts Was Obtained

The reports came to us from each school, listing the number of correct responses, the number of incorrect responses, and the number not attempted, for each question. We were therefore able to determine the percentage that the correct responses were of the total attempts. As indicated above, we threw out of consideration any school in which the "not attempted" was larger for any questions at the end than for questions at the beginning of the test, since this indicated that insufficient time had been given to adequately measure the responses of the children

to the material itself. This "percentage correct" for each book constitutes the third column in Table 1. We have here a picture of the relative ease with which twelve-year-olds read the ten volumes in question. Table 4 gives these percentages by series.

As it is somewhat difficult to interpret the meaning of these differences, we have translated them into a more readily understood medium. Table 5 shows the deviations in terms of Standard Deviation and the average age necessary to read the texts. The percentage of correctness was translated into $\frac{x}{S.D.}$, deviation in terms of the variability of twelve-year-olds. This deviation was translated into a scale ranging from five standard deviations below the mean to five standard deviations above the mean,—calling the former zero, and the latter 100. This places the average at 50. We have values now for each book which correspond to McCall's T-score on the Thorndike-McCall Reading Scale. In Table 5, B-1 is as hard as 43 on the Thorndike-McCall Reading Scale. C-1 is as hard as 45 on the Thorndike-McCall Reading Scale, etc.

Whereas these terms are indispensable statistically to arrive at a final simplified report, they are in themselves so intricate that it is very important to present our relative values in a manner more easily understood. Once arrived at a T-score, it is a simple

matter to use the Thorndike-McCall age norms to find what age is necessary to comprehend any one book. B-1 is as hard as 43 on the reading scale, and the average age of the T-score of 43 is, according to the McCall norms, 130 months. We can therefore, say that children must be this old in order to understand B-1. So with the other reading ages in Table 5. It is also of value to us to know which is the lowest grade able to read each book. These values are derived through the grade norms published by McCall for T-score of his test. We now have before us in Table 5 the average age necessary to read each book and the average grade necessary to read each book.

Selecting Texts to Meet the Needs of Particular Grades and Students

One other immediate advantage, other than the selection of texts, should be parenthetically answered here. By test the mental age of grades used in this study was determined. Thus we could see which grades in any one school are like a fourth grade normally would be; which grades are like a fifth grade normally would be, etc. In this way individual needs of each grade of this city's schools can be seen at a glance. As these scores are on a verbal intelligence test, the differences are in exactly the same ability upon which the relations between texts are based. The correla-

tion between the Thorndike-McCall and the Wylie is high.

A-1 might very well be the best book in every other respect than comprehension. In fact, its very values may be associated with a necessary sacrifice in terms of comprehension. In that case it might be very advisable to use A-1 in all schools where they are able to comprehend it, and to use the book which may not be as valuable, but easier to read, only in those schools where it is necessary to sacrifice the other values in favor of the sine qua non, comprehension. The statement that A-1 is better in other ways is of course entirely hypothetical.

The Variability of Reading Difficulty Within a Single Text

There is a further characteristic of geography text books which our data will allow us to analyze. The difficulty of a book as a book is one major consideration in the treatment of comprehension value, but the degree to which the difficulty varies from very easy to very hard is a very important supplementary consideration. B-1 is easier than any of the other volumes-1 to read. However, if this average ease of comprehension were the result of the product moment of very easy and very hard passages, it would nevertheless be an undesirable text. We want a book the difficulty of which is reasonably constant throughout.

Tables 6 and 7 present the T-scores and reading ages of each question. The reading age is to be interpreted as the average age necessary to read the passage and answer the questions as per the conditions of our test. Some of the disparity in ages is of course due to the sampling of passages and the chance difficulty of the questions, but the range of difficulty of the text book is certainly pictured in a rough manner by the range in difficulties of these passages.

It is especially of note that the difficulty of the sentences in D-1 and D-2 marches progressively from easy to hard, which fact is already noted in connection with Table 1, where the first halves of these two books were shown to be easier than the second halves. The questions are in serial order, and samplings are from the beginning to the end of the book. Whether or not this feature is desirable is dependent upon the progress which children can make in ability to read. The difference in difficulty between the end and beginning of the D's is not justified by the progress which children make in the time covered by the books.

Table 8 gives the standard deviation of reading ages of the questions of each book and the average reading age of the book. It is because of dropping of decimals that the average age of questions per book does not check with the age necessary to read the book,—computed directly. However, the order of the books is of course the same. A consideration of these

measures of variability of the difficulty of passages from the various texts expressed in terms of age shows that as a tendency, the first and second volumes of a series have variabilities very much alike. Note that B-2 and B-1 are together; note that D-1 and D-2 are together; note that A-1 and A-2 are very close together; note that E-1 and E-2 are fairly close together. C-2 and C-1 are very different in variability, but all the other series have Standard Deviations of the volumes-1 which show a likeness to the Standard Deviations of volumes-2. The highest Standard Deviations are for A-2, C-2 and A-1. These three volumes are among the four most difficult to read (Table 5). The lowest standard deviations are those for C-1, E-2, B-2 and B-1. All of these, except E-2 are among the five books which are easiest to read (Table 5).

There is then, luckily, an association of comprehension virtues in these texts. Those that are hard to read are also spotty in their difficulty and those that are easy to read have an even tenor of difficulty.

The Success of the Texts in Reaching Their Reading Public

Table 9 gives the frequency of each reading age of all questions on all volumes-1. An analysis of this frequency distribution of average age necessary to answer each of 104 questions, after reading the

passages selected from five geography text books meant for beginners in geography in grades 4 and 5, will allow us to judge how far the authors of the texts have reached their public. Their public is, of course, the children themselves,—not what teachers think of those children. Although the sales today may be largely based on what superintendents and teachers think children of the fourth and fifth grade can read, rather than what they actually do read, it is just such data as these which will change that emphasis and allow us to judge text books upon a basis of their real ability to get across to the children that they are meant for. The average of the average age necessary to read these questions is 135.79 months. With a standard deviation of about a year and a quarter, about five-sixths of the material is too hard for children in grades 4 to understand. Since one standard deviation below the mean of the distribution, represented in Table 9, is at about 120 months, if we count the children in grades 4 as nearly all younger than 120 months in reading ability, five-sixths of the passages of the texts given them are too hard for the group.

Volumes-2 are a little more adequate to the task which they pretend to perform. The average of the average age necessary to read the 111 questions on volumes-2 is 143.50 months. The standard deviation of reading age of questions is again about a year and a quarter. This distribution of reading ages by

questions on volumes-2 will be found in Table 9. If volumes-2 are meant primarily for grades 6 and 7, only about one-half can be understood by the children of the grade in which they are first introduced, (Grade 6.) Since the average age necessary is about twelve years, if we count most sixth-graders as not more than twelve years old in reading ability they are then able to understand only one-half of the matter presented to them in all geography texts meant for the grade. This one-half is not the first halves of the book as is apparent from Table 7.

Individual Differences in Needs of Various Grades

On January 14, 1921, the 5-A grades of the schools used in this research, about to enter on their 6th grade work had average scores on the Wylie A, B and C ranging from 55.15 to 19.69. Interpreted, this means they range from an average age of 12 years 6 months to 8 years 6 months, in the ability measured by the Wylie test. They differ in the ability of work which they can do as measured by the Wylie test from average 7th grade capacity at beginning of B work to average 4th grade capacity (at point of promotion from B to A).

Associating these facts with those of Table 5, it is apparent that some 6th grades are able to use any of the texts, whereas others need to choose the very

easiest and then still will have a book which is too difficult. The *best* 6th grades can just read the hardest volume-2 and not all can read the easiest. This is a very valuable practical consideration since we may be able to determine as suggested above that a certain book excels the others in all qualities save ease of reading. Then we must know in which grades the difficulty of comprehension makes no difference in order to benefit where we can by the excellence of the book in other qualities.

In order to find which 6th grades (and for volumes-1 which 4th grades) are able to use a difficult book, consult the mental age ratings of each grade.

The Text Already in Use Had Negligible Advantage

In conclusion a word regarding the geography now in use in the lower grades is necessary to assure us that training in it has not prejudiced our results in favor of one of the books. "B," First Book, Part One has been used the last three years. Since the first 90 pages of the 256 (exclusive of indices) of our B-1 is this same Part One, we must be sure that questions on this portion are at least as hard as questions on the remainder of B-1. This portion that has been studied covers questions on our test 1 to 5 inclusive. Though three of these five questions were comparatively easy for our 12-year olds (see Table 6) still Table 1 shows

that the second half of B-1 (which was not studied) was easier than the first half (part of which was studied). We feel therefore that no important advantage was given the "B" series.

This treatise is of comprehension value of the texts only. It makes no claims to ascertain any of the other necessary virtues of a text. When those too are credited through objective means we will be able to choose texts, not risk them.

TABLE 1

Key number of text	Percentage of questions on first half of test answered correctly by a normal group of 12-year-olds	Percentage of questions on second half of test answered correctly by a normal group of 12-year-olds	Percentage of questions on total of test answered correctly by a normal group of 12-year-olds
B-1	68	77	72
C-1	71	67	69
D-1	79	54	66
E-1	64	68	66
B-2	63	64	64
E-2	62	65	64
A-1	62	61	62
D-2	64	51	58
A-2	52	59	56
C-2	54	48	51

TABLE 2.—DISTRIBUTIONS IN WYLIE A, B AND C, JANUARY 14, 1921 BY GROUPS OF SCHOOLS TAKING A, B, C, D AND E GEOGRAPHY READING TESTS-ELEVEN-YEAR-OLDS IN THOSE SCHOOLS AT THAT TIME

	A	*B*	*C*	*D*	*E*	
0	19	34	31	41	33	
0– 5	11	29	24	31	35	→10 % of the number below this line in each distribution.
5– 10	22	21	18	35	25	
10– 15	6	9	4	10	12	
15– 20	11	7	14	18	18	
20– 25	9	15	17	18	22	→25 % of the number below this line in each distribution.
25– 30	12	24	17	29	27	
30– 35	20	15	26	35	31	
35– 40	20	18	21	27	31	
40– 45	11	23	25	24	32	
45– 50	20	36	27	36	36	→This line halves the total number in each distribution.
50– 55	26	47	37	46	56	
55– 60	28	35	32	48	47	
60– 65	37	51	35	64	46	→25 % of the number above this line in each distribution.
65– 70	26	44	29	40	38	
70– 75	17	23	28	37	24	→10 % of the number above this line in each distribution.
75– 80	15	14	17	17	17	
80– 85	13	13	11	12	17	
85– 90	8	6	8	10	7	
90– 95	5	7	9	12	7	
95–100	4	7	8	9	6	
100–105	2	4	4	6	3	
105–110	0	1	2	3	0	
110–115	1	1	2	3	0	
115–120	1	2	1	1	2	
120–125	..	..	1	..	1	
	344	486	448	612	573	

TABLE 3.—DISTRIBUTIONS IN WYLIE A, B, AND C, JANUARY 14, 1921, BY GROUPS OF SCHOOLS TAKING A, B, C, D AND E GEOGRAPHY READING TESTS—TWELVE-YEAR-OLDS IN THOSE SCHOOLS AT THAT TIME

	A	*B*	*C*	*D*	*E*
0	18	29	17	28	35
0– 5	11	10	17	27	17
5– 10	11	9	13	23	12
10– 15	4	7	6	12	8
15– 20	6	7	10	16	14
20– 25	3	12	11	16	9
25– 30	7	11	12	17	16
30– 35	14	22	10	23	23
35– 40	14	14	23	25	28
40– 45	26	23	12	34	23
45– 50	19	19	24	37	26
50– 55	37	30	39	51	56
55– 60	37	28	34	46	40
60– 65	35	39	37	58	60
65– 70	30	40	35	50	45
70– 75	21	23	33	33	34
75– 80	16	21	26	31	35
80– 85	20	18	25	32	23
85– 90	13	11	18	15	19
90– 95	9	5	13	13	9
95–100	8	2	7	3	8
100–105	2	4	8	5	6
105–110	0	4	4	8	3
110–115	1	4	4	4	4
115–120	1	0	3	1	0
120–125	..	2	1	3	1
125–130	..	2	1	0	2
130–135	..	0	1	0	..
	363	397	444	612	556

TABLE 4

TEXT SERIES	PERCENTAGE OF QUESTIONS CORRECT
B	67.7
E	64.8
D	62.1
C	59.4
A	58.6

TABLE 5.—DISTINCTIONS IN COMPREHENSION VALUE TRANSLATED INTO AGE AND GRADE

Text	Proportion of correct responses	$\frac{x}{S.D.}$ in a normal distribution of difficulty	T-scale like Thorndike-McCall reading scale	Average age necessary to comprehend	Average grade necessary to comprehend
B-1	.72	−.58	44.2	133	5.48
C-1	.69	−.50	45.0	135	5.50
D-1	.66	−.41	45.9	138	5.66
E-1	.66	−.41	45.9	138	5.66
B-2	.64	−.36	46.4	139	5.74
E-2	.64	−.36	46.4	139	5.74
A-1	.62	−.31	46.9	141	5.82
D-2	.58	−.20	48.0	144	6.00
A-2	.56	−.45	48.5	145	6.09
C-2	.51	−.02	49.8	150	6.31
Whole series				Note columns above for age and grade per volume	
B	.68	−.47	45.3	135	5.51
E	.65	−.39	46.1	138	5.69
D	.62	−.31	46.9	141	5.82
C	.59	−.23	47.7	143	5.95
A	.59	−.23	47.7	143	5.95

TABLE 6.—T-SCORES AND READING AGES BY QUESTIONS IN TESTS ON VOLUMES-1

Number of question	T-score	Reading age	T-score	Reading age	T-score	Reading age	T-score	Reading age	T-score	Reading age
	B-1		C-1		D-1		E-1		A-1	
1	40.1	121	48.2	145	39.2	119	34.4	105	42.6	129
2	35.2	108	46.7	140	39.2	119	41.2	125	40.8	123
3	45.3	136	38.2	117	33.5	102	52.8	157	50.3	151
4	46.7	140	43.3	131	39.6	120	44.5	135	35.9	110
5	35.9	110	42.9	130	40.5	123	46.4	139	47.7	143
6	45.0	135	43.6	132	43.6	132	37.7	115	44.5	134
7	49.5	148	51.0	152	38.7	118	54.7	163	46.1	138
8	46.1	138	42.6	129	46.9	141	51.8	154	48.2	145
9	56.7	169	40.5	122	43.6	132	46.7	140	54.1	161
10	42.3	128	40.8	123	44.5	134	45.9	138	61.3	183
11	36.6	112	45.3	136	42.9	130	37.2	114	42.6	129
12	45.3	136	42.9	130	45.3	136	49.5	148	42.9	130
13	49.2	148	45.3	136	48.2	145	40.8	123	46.9	141
14	38.2	117	44.2	134	54.1	161	46.9	141	46.7	140
15	39.2	119	51.0	152	45.6	137	45.9	138	48.7	146
16	42.6	129	43.3	131	38.7	118	48.0	144	54.7	163
17	44.5	135	42.3	128	45.6	137	42.6	129	47.5	143
18	42.6	129	48.5	146	47.5	143	49.5	149	56.7	169
19	42.9	130	47.7	143	56.7	169	44.2	134	52.0	155
20	....	...	....	...	39.2	119	....	...	36.6	112
21	....	...	....	...	51.5	153	....	...	42.0	127
22	....	...	....	...	56.1	167	....	...	41.6	126
23	....	...	....	...	50.8	152	....	...	42.9	130
24	....	...	....	...	50.5	151				

TABLE 7.—T-SCORES AND READING AGES BY QUESTIONS IN TESTS ON VOLUMES-2

Number of question	B-2		E-2		D-2		A-2		C-2	
	T-score	Reading age	T-score	Reading age	T-score	Reading age	T-score	Reading age	T-score	Reading age
1	40.1	121	52.0	155	45.3	136	57.1	169	52.8	157
2	53.3	159	47.5	142	41.6	126	55.5	166	44.8	135
3	45.0	135	46.9	141	45.9	138	59.2	176	46.9	141
4	47.5	143	44.8	135	45.9	138	58.0	172	42.0	127
5	47.7	144	42.6	129	51.3	153	58.4	173	46.4	139
6	47.7	144	51.0	152	53.9	161	59.2	176	55.8	166
7	50.8	152	41.6	126	45.3	136	46.1	138	54.7	163
8	45.3	136	41.2	125	37.7	115	49.2	148	56.1	167
9	46.1	138	53.3	159	40.8	123	41.2	125	37.2	114
10	42.3	128	42.0	127	43.9	133	49.0	147	46.7	140
11	44.2	134	40.5	122	47.7	143	42.0	127	52.5	156
12	49.7	149	51.8	154	51.8	154	42.0	127	57.7	171
13	52.0	155	52.3	156	47.2	142	35.9	110	57.1	169
14	45.9	138	44.5	135	53.1	158	41.6	126	50.8	152
15	54.4	162	43.9	133	53.6	160	41.2	125	47.2	142
16	38.2	117	43.9	133	47.5	142	57.1	169	36.6	112
17	37.7	115	45.6	137	51.0	152	49.5	148	55.8	166
18	45.3	136	47.7	143	59.5	177	51.0	152	51.5	154
19	46.7	140	....	...	43.6	132	56.1	167	38.7	118
20	43.6	132	....	...	45.9	138	48.7	146	50.0	150
21	45.9	138	....	...	42.9	130	45.3	136	53.9	161
22	....	...	....	...	43.6	132	47.7	143	53.3	159
23	....	...	....	...	51.5	153	42.6	129		
24	....	...	....	...	60.4	179	48.0	144		
25	....	...	....	...	....	...	45.3	136		
26	....	...	....	...	....	...	40.5	123		

TABLE 8.—STANDARD DEVIATIONS AND AVERAGES OF READING AGES OF THE QUESTIONS BY BOOK IN ORDER OF SIZE OF STANDARD DEVIATIONS

Book	S.D.	M
C-1	9.45	134.58
E-2	11.48	139.11
B-2	11.79	138.86
B-1	14.55	130.95
E-1	14.64	136.37
D-2	15.56	143.79
D-1	16.79	135.75
A-1	17.32	140.35
C-2	17.66	148.14
A-2	18.79	146.08

TABLE 9.—DISTRIBUTION OF READING AGE NECESSARY TO READ QUESTIONS ON VOLUMES-1 AND VOLUMES-2

Reading age	Volumes-1	Volumes-2
100–105	1	
105–110	2	
110–115	5	3
115–120	9	4
120–125	7	4
125–130	11	13
130–135	15	8
135–140	15	19
140–145	12	15
145–150	9	5
150–155	8	11
155–160	1	9
160–165	4	5
165–170	4	8
170–175	..	3
175–180	..	4
180–185	1	
	Volumes-1	Volumes-2
N	104	111
M	136.20	143.95
S.D.	15.25	16.05

15.
EP 1
3 1r

LB3045 .F7
Textbook selection,
Gutman Library
APS7461

3 2044 028 935 68

DATE DUE

DEC 04 1994

JA 3 1994

DEMCO 38-297